Analysis of Drum Machine Kick and Snare Sounds

Jordie Shier, Kirk McNally and George Tzanetakis

University of Victoria

Introduction and Motivation

- The use of electronic drum samples is widespread in contemporary music productions, particularly in Electronic Dance Music (EDM).
- Intelligent Music Production (IMP), including creating tools to assist users with the organization and management tasks associated with libraries of this type is a growing area of research.[1]
- The study here explores the use of time segmentation methods to better characterize kick and snare drum samples, the results being of use for the development of IMP tools.
- A database of 4230 kick and snare samples, representing 250 individual electronic drum machines was used in this study.

Methodology

- Analysis of kick and snare drum sounds using a time segmentation scheme informed by prior work into the characterization of percussive sounds. [1-5].
- Audio feature extraction performed using the Essentia library [6].
- Audio features include: Bark bands, MFCCs, HFC, spectral and temporal features (133-dimension feature space).
- Principle Component Analysis (PCA) is used to reduce the dimensionality of the original feature space and to examine the effect that time segmentation has on feature variance in lower dimensions.

Experimental Results

- Feature variance is maximized in the first two dimensions when using a 100ms time segment starting at 20% of the attack for kick sounds, and a 250ms time segment starting at 50% of the attack for snare sounds.
- Figures 1 & 2 show plots of notable drum machines after PCA for kicks and snares respectively.
- Table 1 shows PCA variance ratios for both kicks and snares.
- Audio classification using Scikit-learn [7] is used to compare the effect of the time segmentation.
- Using a mixed time segmentation scheme resulted in the best classification scores for both kicks and snares, reporting 84.20% and 69.88% accuracy respectively.

[1] AES Workshop on Intelligent Music Production (WIMP) 2015, 2016, 2017

Figure 1: Plot of kicks after PCA using mixed time segmentation

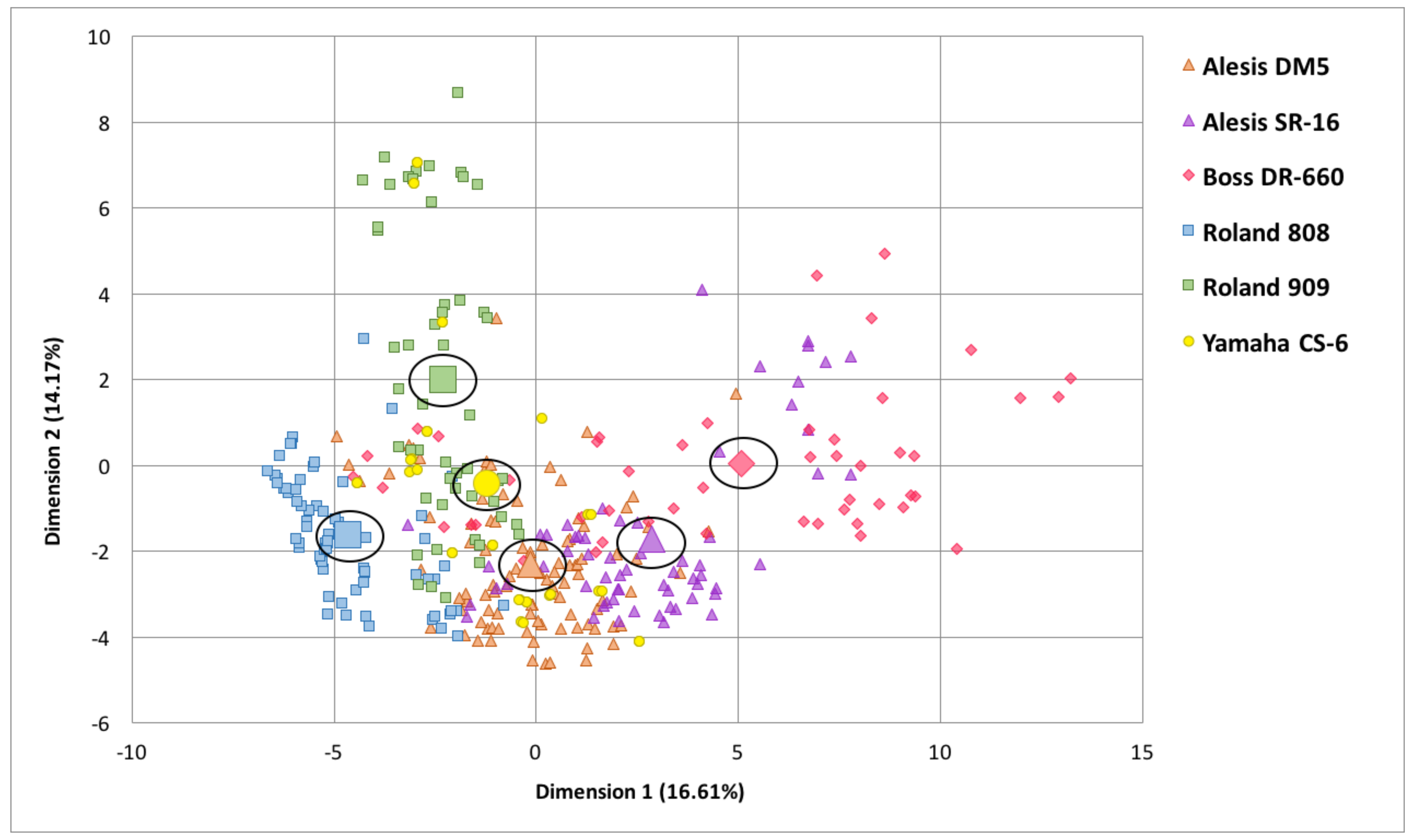

Figure 2: Plot of snares after PCA using mixed time segmentation

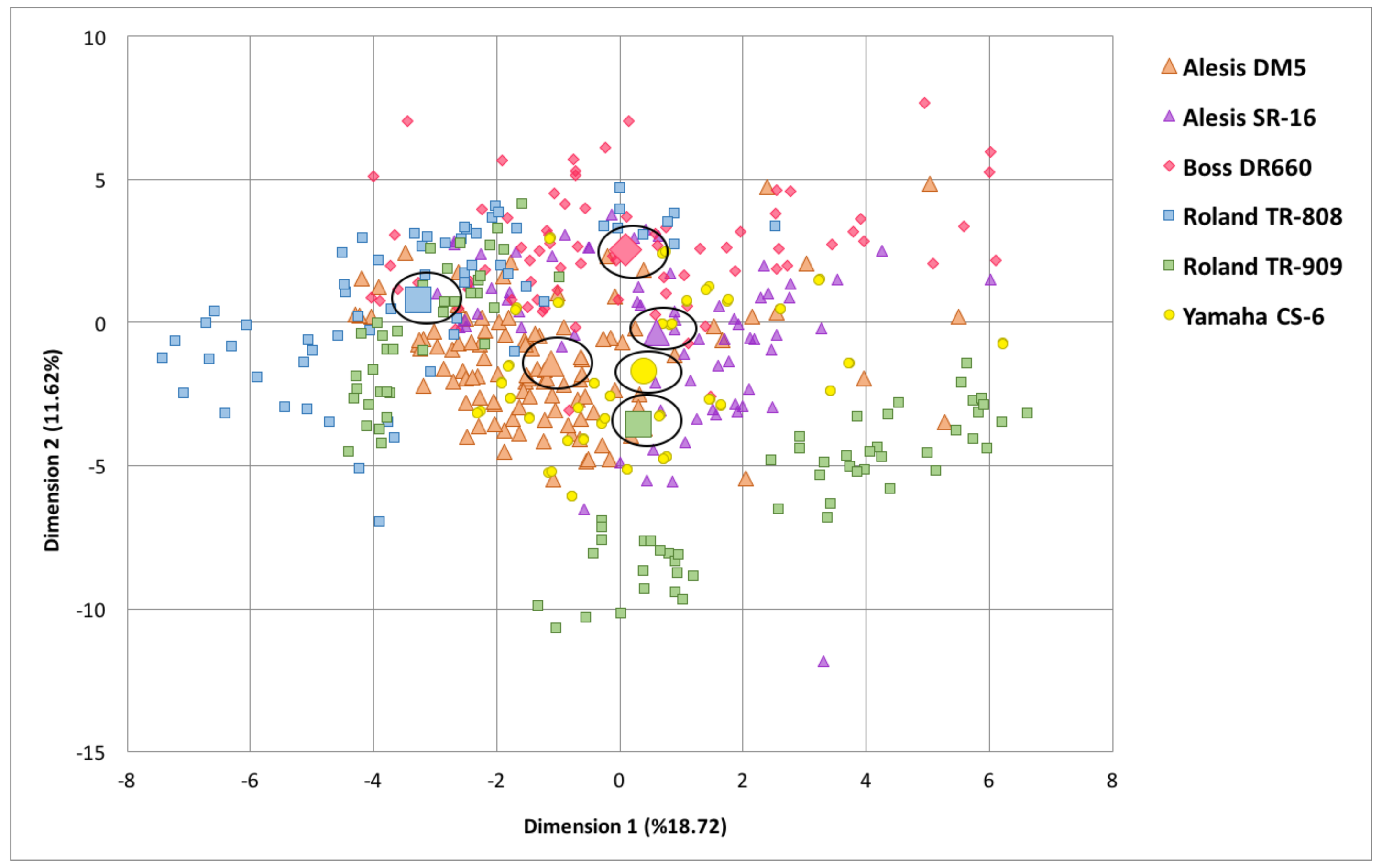

Table 1: Variance ratios from PCA after applying various time segmentations

Length	Start	Kick					Snare				
		Dim 1	Dim 2	Dim 3	Dim 4	1+2	Dim 1	Dim 2	Dim 3	Dim 4	1+2
$25ms$	20%	17.95%	12.46%	9.44%	7.91%	30.41%	17.11%	13.85%[5]	9.51%	5.48%	30.97%
$25ms$	50%	17.64%	11.93%	9.52%	7.65%	29.57%	18.08%	13.73%	9.39%	5.37%	31.81%
$25ms$	90%	15.69%	11.53%	9.60%	7.76%	27.21%	19.38%	12.81%	8.79%	5.36%	32.18%
$100ms$	20%	17.30%	14.35%	9.43%	7.80%	31.65%[3]	20.16%	11.03%	9.85%	5.93%	31.19%
$100ms$	50%	16.72%	13.65%	9.41%	8.22%	30.37%	20.75%	10.75%	9.77%	5.95%	31.32%
$100ms$	90%	15.62%	12.45%	10.57%	8.70%	28.08%	21.22%	10.5%	9.37%	6.26%	31.37%
$250ms$	20%	17.01%	14.40%[2]	8.95%	8.18%	31.41%	21.22%	10.73%	9.22%	6.52%	31.95%
$250ms$	50%	16.48%	13.83%	9.16%	8.16%	30.30%	21.70%	10.54%	9.08%	6.52%	32.24%
$250ms$	90%	15.31%	12.92%	9.91%	8.79%	28.23%	22.75%[4]	10.04%	8.89%	6.63%	32.79%[6]
$500ms$	20%	17.16%	13.52%	8.87%	7.83%	30.68%	21.43%	10.19%	9.15%	6.94%	31.62%
$500ms$	50%	16.52%	13.10%	8.92%	8.06%	29.63%	21.86%	10.01%	9.03%	6.97%	31.87%
$500ms$	90%	15.16%	12.70%	9.47%	8.71%	27.86%	22.70%	9.69%	8.71%	7.01%	32.39%
Full	0%	18.03%[1]	13.44%	9.07%	7.37%	31.47%	21.01%	10.38%	9.15%	7.08%	31.39%

Main contributing features:

[1] **Dim 1:** HFC, HFC Std Dev, Mid-High Spectral Energyband
[2] **Dim 2:** Spectral Flatness dB, Spectral Centroid, Spectral Kurtosis
[3] **Dim 1:** HFC, HFC Std Dev, High Spectral Energyband; **Dim 2:** MFCC Band 2, Mid-Low Spectral Energyband Std Dev, Mid Low Spectral Energyband
[4] **Dim 1:** Spectral Energy, Bark Band 18 and 19
[5] **Dim 2:** Spectral Decrease, Spectral Decrease Std Dev, Spectral RMS
[6] **Dim 1**:Spectral Energy, Bark Band 18 and 19 **Dim 2**: Bark Spread Std Dev, Zero Crossing Rate Std Dev, MFCC Band 5

Conclusion and Future Work

- Results show that time segmentation effects the variance and distribution of each audio feature in a unique way.
- Using a mixed time segmentation approach leads to improved classification results and the ability to retain variance in lower dimensions when using PCA.
- Future work includes the perceptual testing of the techniques described here in order to determine whether utilizing time segmentation methods in pre-processing leads to more perceptually relevant characterization of audio samples.

References

- [1] Herrera, P., Yeterian, A., and Gouyon, F., “Automatic classification of drum sounds: a comparison of feature selection methods and classification techniques,” in Music and Artificial Intelligence , pp. 69–80, Springer, 2002.
- [2] Herrera, P., Dehamel, A., and Gouyon, F., “Automatic labeling of unpitched percussion sounds,” in Audio Engineering Society Convention 114 , Audio Engineering Society, 2003.
- [3] Pampalk, E., Herrera, P., and Goto, M., “Computational models of similarity for drum samples,” IEEE transactions on audio, speech, and language processing , 16(2), pp. 408–423, 2008.
- [4] Danielsen, A., Waadeland, C. H., Sundt, H. G., and Witek, M. A., “Effects of instructed timing and tempo on snare drum sound in drum kit performance,” The Journal of the Acoustical Society of America , 138(4), pp. 2301–2316, 2015.
- [5] Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., and Sandler, M. B., “A tutorial on onset detection in music signals,” IEEE Transactions on speech and audio processing , 13(5), pp. 1035–1047, 2005.
- [6] D. Bogdanov, N. Wack, E. G«omez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, “Essentia: an open-source library for sound and music analysis,” in 21st ACM International Conference on Multimedia , pp. 855–858, ACM, 2013.
- [7] F. Pedregosa et al. , “Scikit-learn: Machine learning in python,” Journal of Machine Learning Research, vol. 12, no. Oct, pp. 2825–2830, 2011.